DOCUMENT RESUME

ED 407 671 CS 215 797

AUTHOR Reinwein, Joachim; Huberdeau, Lucie

TITLE A Second Look at Dwyer's Studies by Means of Meta-Analysis: The Effects of Pictorial Realism on Text Comprehension and Vocabulary.

PUB DATE Apr 97
NOTE 35p.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Factor Analysis; Foreign Countries; Higher Education; *Illustrations; Meta Analysis; *Reader Text Relationship; *Reading Comprehension; Reading Research; Secondary Education; *Vocabulary
IDENTIFIERS Dwyer (Francis M); *Text Processing (Reading)

ABSTRACT

A meta-analysis examined a series of studies by F.M. Dwyer on the effect of illustrations on text comprehension. Principal component analysis was used to reduce the four posttests used by Dwyer to more fundamental factors of learning, followed by analyses of variance. All nine studies (involving secondary-school and college students) in which Dwyer provides the mean results for each experimental group and its control group for the identification, drawing, comprehension, and terminology tests were used. Results indicated that 2 factors (a vocabulary learning factor and a text-comprehension factor) accounted for most of the variance. Results also indicated a profound dichotomy between the vocabulary learning factor and the text comprehension factor results--the absence and presence of pictures, the degree of pictorial realism, and the absence and presence of color are all significant variables with respect to the vocabulary factor, but nonsignificant with respect to the text-comprehension factor. Findings suggest significant main effects only for the vocabulary factor, thus putting into question Dwyer's central hypothesis that realistic pictures accompanying text are significantly less effective than abstract ones. (Contains 87 references, 13 tables and 1 figure of data. Appendixes provide examples from the four posttests used in Dwyer's studies.) (RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 407 671

A second look at Dwyer's studies by means of meta-analysis: the effects of pictorial realism on text comprehension and vocabulary

Joachim Reinwein
Université du Québec à Montréal, Canada

Lucie Huberdeau
Gouvernement du Québec, Canada

April 1997

Illustration's effect on text comprehension has been the subject of numerous empirical studies (see the reviews of Readence & Moore, 1981; Levie & Lentz, 1982; Goldsmith, 1984; Houghton & Willows, 1987; Willows & Houghton, 1987; Reinwein, forthcoming; note 1). Some of these studies focussed more specifically on the effect of pictorial realism on text, oral or written. Certain researchers, and specially Dwyer, consider pictorial realism to be representable as a continuum, with the most true-to-life illustrations at one end (i.e. color photographs) and simple line drawings in black and white at the other.

Dwyer's studies figure prominently among picto-verbal research, as much chronologically as quantitatively (see Table 2). To be added to this list are doctoral theses that have been written under his guidance (Wheelbarger, 1970; Parkhurst, 1974; Joseph, 1978; de Melo, 1980; Lamberski, 1980) and the articles he has written in collaboration with others (Lamberski & Dwyer, 1983; Parkhurst & Dwyer, 1983).

Dwyer's studies also stand out in picto-verbal research for another reason: The studies compare as many as eight levels of pictorial realism within each study. Starting in 1967, Dwyer spent almost two decades examining the effect of pictorial realism on different comprehension and vocabulary measures. Sometimes he used attitudinal measures and, as a control measure, study time. In his experiments, Dwyer followed essentially the same experimental procedure, used the same experimental material and tested his hypotheses using the same basic measures throughout. His research included up to nine experimental versions analyzed by means of five dependent comprehension and vocabulary measures coming from four posttests, including in some cases repeated measures on the delay variable (immediate vs. delayed posttests). In these studies, there were up to 360 binary statistical comparisons, presented sometimes by means of 10 separate one-way analyses of variance. As a consequence, the implications of these comparisons are not always easy to understand. The situation becomes even more complicated when trying to do a cross-studies synthesis of results. Dwyer's (1972b, 1978) own across-studies synthesis was limited to a mainly qualitative-interpretative approach. This should not be so. We think that statistical across-studies synthesis by means of meta-analysis is possible -- even desirable. It is our contention that given the homogeneous character of his experiments and the abundance of measures within each of them, meta-analysis of his data can be done using classical methods of statistical synthesis, i.e. factor analysis.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. Reinwein
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

AS 215797

Pictorial realism and color in experimental research

Before studying Dwyer's own experiments, let us review the pictorial-realism literature in its broader sense. Indeed, as can be seen by the diversity of the pictorial opposition studied in picto-verbal text research (e.g., literal vs. analogous pictures, pictures representing details vs. main ideas, two-dimensional vs. stereogram pictures), the concept of pictorial realism is not limited to a single dimension. Some authors, Dwyer, among others, consider color to be a part of pictorial realism (i.e. black & white vs. colored pictures). Note that the following categorization of the pictorial-realism literature is being done without prejudice to this integrative point of view:

(a) Research which obtained nonsignificant main effects concerning the pictorial-realism variable (color excluded): Most studies were of this type. In Koenke & Otto (1969), text passages were accompanied by a picture specifically or generally relevant to the main idea. The specifically relevant picture depicted the essence of the main idea whereas the generally relevant picture illustrated the general content of the passage. In Haring & Fry (1979), a prose passage was accompanied by pictures depicting either the main ideas or both the main ideas and the nonessential details. Burdick (1959) and Reid, Briggs & Beveridge (1983) compared cross-sectional and three-dimensional pictures. Smith & Smith (1991) compared abstract and concrete illustrations: abstract illustrations represented a given concept through the use of circles, arrows and rectangles whereas concrete illustrations represented "the best example" of the concept. Denis & Pouqueville (1976) compared photographs, a film and drawings; Thomas (1978), compared color photographs to line drawings; Jagodzinska (1976), compared so-called schematic to realistic illustrations. In Koski (1975), finally, an experimental version deemed "costly" (i.e. using realistic color illustrations) was prepared by a professional using the most sophisticated production techniques. It was compared to a low-cost experimental version (i.e. less-realistic illustrations which focussed only on aspects deemed essential which was prepared by a primary teacher using only basic production techniques.) In all of these studies, results defied expectations: the varying illustrated experiment versions did not produce significant results.

(b) Research which obtained significant interactional effects concerning the pictorial-realism variable (color excluded): In three studies (Hurt, 1987; Waddill, McDaniel & Einstein, 1988; McAlister, 1991), there was a significant interaction between the pictorial variable and the type of response elicited by the test. Hurt (1987) compared literal and analogous illustrations to concrete and abstract text information, respectively called phenomenological and non-phenomenological information. The interactional effect indicated that literal illustrations favored the recall of phenomenological information while non-phenomenological information was better remembered using analogous illustrations. Waddill, McDaniel & Einstein (1988) compared the effect of line drawings. The line drawings illustrated either details mentioned in single propositions or the more important "relational" ideas on free and guided recall (i.e. ideas presented in different propositions). Here, once again, there was a significant interaction between the illustration-type and the information-type recalled: in some of the comparisons, the recall of linguistic details was higher after exposure to illustrations representing details, whereas the recall of linguistic main ideas benefited from illustrations of these main ideas. In McAlister (1991), the effectiveness of

the two types of illustrations, analogous and schematic, was also found to depend on the recall-type (physical characteristics of the described objects vs. the abstract relationships).

(c) Research which obtained significant main effects concerning the pictorial-realism variable (color excluded): There was only one study (i.e. Hannafin, 1983) where the main effect of the pictorial variable was significant and this, without interaction with other independent variables. In Hannafin (1983), an audiotape presentation of a short story was accompanied by either global pictures representing target- and non-target information (the so-called SIMPLE condition) or by global and "close-up" pictures (called the CLOSE-UP condition). The close-up pictures focussed on target information. This second experimental version was shown to be superior for the recall of abstract and concrete target information.

(d) Research with color as an independent variable: Since Stroop (1935), color had been used on more than one occasion to create semantic interference. Color has also been used as a mnemonic device, both for linguistic stimuli (Hinds, 1966; Gattegno, 1966; Taber & Glaser, 1962; Giasson-Lachance & Leduc, 1981; Glaser & Glaser, 1982) and pictorial stimuli. In the latter case, the effect of color was studied with respect to affective measures (e.g., Bloomer, 1960; Amsden, 1960; Sloan, 1971; Samuels, Biesbrock & Terry, 1974; Booth & Miller, 1974; Ramsey, 1982), perceptual measures (Fleming & Sheikhian, 1972; Franzwa, 1973; Bohle & Garcia, 1987) and cognitive measures (see reviews by Lamberski & Roberts, 1979 and by Dwyer & Lamberski, 1982, among others). Among the latter studies, Katzman & Nyenhuis' (1972) and Chute's (1980) indicated that, compared to black and white illustrations, color illustrations improved the recall of accessory details, whereas the recall of main ideas was not affected. In Reid, Briggs & Beveridge (1983), finally, the color variable proved to be nonsignificant.

Description of Dwyer's studies

In his studies on pictorial realism, Dwyer varied the pictorial material according to two parameters, i.e. the degree of pictorial realism from simple line drawings to realistic photographs and the presence and absence of color. (Factorial combination of these parameters allow distinctions to be drawn between pictorial realism and color-induced effects and the realism-color interaction to be studied, this was not possible at the time where Dwyer used one-way analyses of variance.)

Dwyer's linguistic material consisted of a text of approximately 2000 words describing the parts and functions of the human heart as well as about forty illustrations (or "visuals") each corresponding to a passage. Depending on the experiment in question, the text was accompanied either by three (ex. Dwyer 1968c), four (ex. Dwyer, 1968e) or eight different illustrated experimental versions (ex. Dwyer 1967c), all of which were normally compared to a non-illustrated experimental version. In Table 1, we have indicated the nature of the nine versions used in Dwyer (1967c) which span the entire pictorial continuum (see Appendix A for a sample of the visuals used):

Table 1

Dwyer's control and experimental versions (e.g., 1967c)

Version 1: Text without visuals of the heart (control)
Version 2: Text and simple line drawings of the heart (black & white)
Version 3: Text and simple line drawings of the heart (coloured)
Version 4: Text and detailed, shaded drawings of the heart (black & white)
Version 5: Text and detailed, shaded drawings of the heart (coloured)
Version 6: Text and photographs of a heart model (black & white)
Version 7: Text and photographs of a heart model (coloured)
Version 8: Text and realistic photographs of the heart (black & white)
Version 9: Text and realistic photographs of the heart (coloured)

The text was presented either in listening mode (ex. 1967a) or in reading mode (ex. 1967b). In the listening mode, the text was delivered via a recording and the illustrations were projected onto a screen using slides. The subjects were thus obliged to follow at a set rhythm. In the reading mode, the subjects worked at their own pace as both the text and the illustrations were provided in the form of a booklet. In some studies (ex. 1972a), supplementary comprehension questions were inserted at strategic points in the text in order to focus subject's attention on the data. In all of Dwyer's studies, four posttests identified by Dwyer as drawing, identification, terminology and comprehension tests were administered immediately after the text presentation. In one of them (1967c), the tests were repeated a second time one month later:

- Identification test: This test was intended to measure S's ability to identify numbered parts on a detailed, shaded drawing of the heart (see Appendix B).

- Terminology test: This test was intended to evaluate knowledge of referents for specific symbols (see Appendix C).

- Drawing test: This test was intended to evaluate S's learning of specific locations of the parts of the heart (see Appendix D).

- Comprehension test: This test was intended to measure understanding of the heart (its parts and its internal functions) (see Appendix E).

According to Dwyer, the four tests corresponded to four distinct learning objectives. They provided Dwyer with five measures, the fifth being the total score of the four previous measures.

Table 2 summarizes the main characteristics of Dwyer's studies.

Table 2
Main characteristics of Dwyer's studies

ARTICLE	SUJBECTS	EXPERIMENTAL VERSIONS						4 POSTTESTS
Year of publi-cation	Degree	Total	Type of visual*	Black&W Colour(BW) / Color (C)	Nonill. version	Oral / written text	Questio ns inserte d in text	Immediate (I) / delayed (D)
1967a*	university	4	S, D, R	BW	yes	oral	no	I
1967b*	university	4	S, D, R	BW	yes	written	no	I
1967c*	sec. 9-12	9	S, D, M, R	BW, C	yes	oral	no	I, D
1968a	sec. 10	9	S, D, M, R	BW, C	yes	oral	no	I, D
1968b	university	9	S, D, M, R	BW, C	yes	written	no	I
1968c	sec. 9	5	S, D, R	BW	yes	written	no	I, D
1968d	sec. 11	9	S, D, M, R	BW, C	yes	oral	no	I, D
1968e*	university	5	S, D, M, R	BW	yes	oral	no	I
1969a	sec. 10	9	S, D, M, R	BW, C	yes	oral	no	I, D
1969b	university	4	S, D, R	BW	yes	written	yes	I
1970a	university	5	S, D, M, R	BW	yes	oral	no	I
1970c*	university	5	S, D, M, R	BW	yes	oral	no	I
1971a	university	9	S, D, M, R	BW, C	yes	written	yes	I
1971b*	university	9	S, D, M, R	BW, C	yes	oral	no	I
1971c	university	9	S, D, M, R	BW, C	yes	written	yes	I
1971d	university	9	S, D, M, R	BW, C	yes	written	no	I
1971e	university	9	S, D, M, R	BW, C	yes	written	yes	I
1971f	university	9	S, D, M, R	BW, C	yes	oral	no	I
1972a*	university	9	S, D, M, R	BW, C	yes	written	yes	I
1975*	university	9	S, D, M, R	BW, C	yes	written	yes	I
1976*	university	8	S, D, M, R	BW, C	no	oral	no	I

* Type of visual: S = simple line drawing; D = detailed, shaded drawing; M = model photograph; R = realistic photograph.

In the interest of group homogeneity, subjects were randomly assigned to the different experimental test versions in most of Dwyer's studies. This homogeneneity was also tested with a so-called physiology pre-test. Once treatment homogeneity was ascertained, Dwyer compared the different experimental versions using a one-way analysis of variance for each of the five dependent measures and post-hoc tests. So it was that with nine versions Dwyer obtained as many as 180 statistical comparisons for immediate testing (36 x 5) and, if applicable, the same amount for delayed testing. Unfortunately, the impressive number of comparisons is a major obstacle to the understanding of what is really meant and it complicates between-study synthesis, even with the addition of a fifth and supposedly synthetic measure (i.e. total score). Note that three of the four posttests are highly correlated (cf. Table 4) and as a consequence, the learning objective associated with the fourth test is misrepresented in the total score.

In order to avoid the two extreme solutions which lead, on the one hand, to an over-abundance of measures and, on the other hand, to a misrepresentative composite measure, we will use principal component analysis, which is a factor

BEST COPY AVAILABLE

analysis, as a statistical synthesis method to reduce the four tests to more fundamental factors of learning, followed by analyses of variance. This methodology will provide generalizable statistical results that reach beyond the limits of the individual studies. In so doing, we get a better overall picture than with an exclusively qualitative approach (Dwyer, 1970b, 1972b, 1978).

Meta-analysis of Dwyer's studies

To begin, we have selected all studies in which Dwyer provides the mean results for each experimental group (i. e. the illustrated-text versions) and its control groups (i.e. the non-illustrated version of the same text) for the identification, drawing, comprehension and terminology tests (1967a, 1967b, 1967c, 1968e, 1970c, 1971b, 1972a, 1975, 1976). These studies are marked with an asterix in Table 2. Dwyer's (1968a, 1968d, 1969a) articles each describe a part of his (1967c) report.

We are particularly interested in the average raw scores on the four tests. With the exception of three studies, Dwyer checked the homogeneity between groups by pre-testing the subjects on previously acquired physiological knowledge. The groups were statistically equivalent. In the three other studies, our own statistical analysis of the previous knowledge of the experimental and control groups reveals that they are indeed statistically equivalent. There is no significant relationship between the results of the physiology pre-test and the degree of realism of the illustrations ($F = 0.55$, $df = 3,27$; $p = 0.65$) or with the presence of color ($F = 0.86$, $df = 2,27$; $p = 0.43$). That is why our analysis is based on the mean raw scores for the four tests provided by the 123 independent experimental groups. Our database actually includes 159 groups, if we take into consideration the fact that the (1967c) study employs delayed posttests which added 36 more measures to those obtained from immediate posttests.

In Dwyer's studies the average number of subjects per group is 27.5 ($SD = 7.65$). Half of the groups have a number of subjects ranging from 22 to 30. Varying between 36 and 62 subjects, the size of the four experimental and control groups from the (1968e) and (1970c) experiments was above average. The smallest groups are found in Dwyer (1976): they ranged between 11 and 33 subjects. Setting aside the two groups made up of 11 and 14 subjects, the average group results were all based on at least fifteen subjects. In accordance with the Central Limit Theorem, which states that averages calculated from more than fifteen observations are normally distributed, the per-group results can, therefore, be considered a normal variable from the very outset of the study.

The characteristics of the 123 independent groups are shown in Table 3. The varying experimental conditions limit which hypotheses can be tested. Accordingly, a comparison of immediate and delayed posttest results can only be done for the secondary school level results. The importance of the mode of text presentation (written, oral) can only be examined with regards to university subjects.

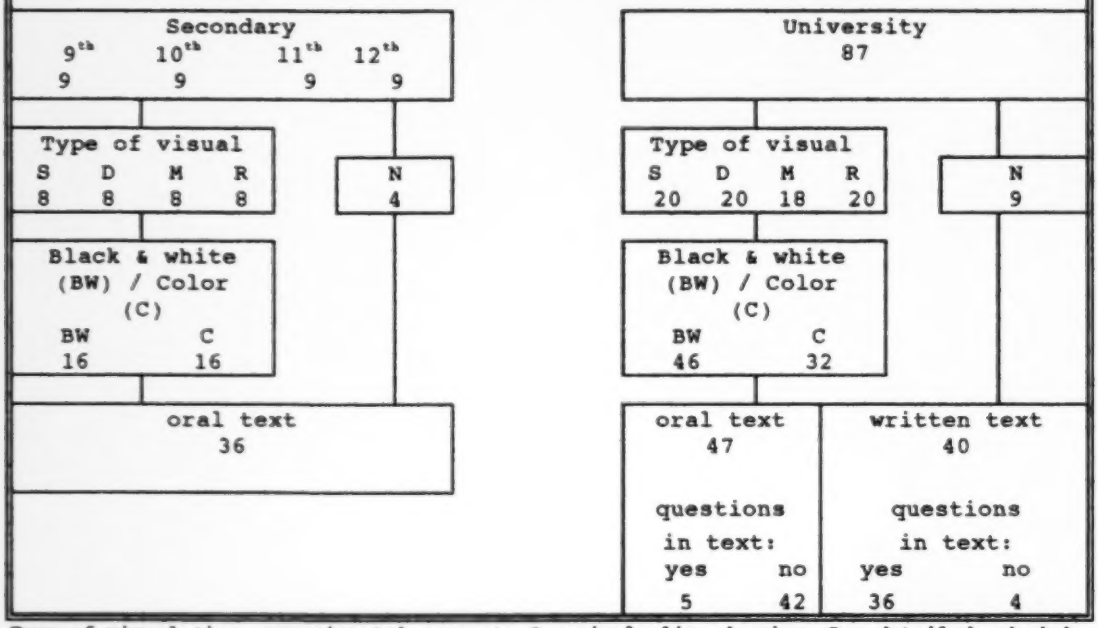
The secondary school results all originate from the same study (Dwyer, 1967c) even if data from that study are also used elsewhere (Dwyer, 1968a, 1968d, 1969a; see Table 2). In one experiment (Dwyer, 1967c) do the subjects take the posttest twice, immediately after the experiment and again one month later. Each of the four secondary school levels has nine groups. The students, in this case, are randomly assigned either a control treatment (N) or one of the eight experimental treatments (S-BW, S-C, D-BW, D-C, M-BW, M-C, R-BW, R-C). All of the subjects listened to the recorded text and examined the illustrations which were projected from slides. This study of 36 independent and twice-tested groups resulted in a total of 72 sets of results per test. It is therefore possible, at least at the secondary school level, to simultaneously evaluate the main effects of Pictorial realism (S, D, M, R), Color (BW, C), Delay between the experimental treatment and the tests (I, D) and their interactions (Pictorial realism x Color, Pictorial realism x Delay, Color x Delay). Further analysis using contrastive variance can then highlight the relations between the four control groups and the 32 experimental groups.

The university results originate from eight studies. In Dwyer's earlier studies, the illustrations were provided only in black and white. As a result, a majority of groups were shown the illustrations in black and white (46 versus 32). Similarly, the level of realism as incarnated by photographed plaster models was less frequent (M = 18) than the other levels of realism (S, D, R = 20 each). Written text (40) is employed slightly less than oral text (47). These distinctions do not however go beyond acceptable limits. More critical is the possibility of confusing the effects of the text-presentation mode with the effects of text-inserted questions: in the reading mode, 36 groups faced inserted questions whereas four did not. In the listening mode, it's the opposite: five groups faced inserted questions while fourty-two did not. Thus, at the university level, it is possible to verify the effects of the level of realism and color as well as the interaction between these two variables, while simultaneously verifying the importance of text-presentation modes on the performance of the 68 experimental groups, again using three-way analysis of variance. Further analysis using contrastive variance can highlight the relationships between the nine control groups and the 68 experimental groups.

BEST COPY AVAILABLE

Table 3

Characterization of the 123 experimental and control groups used in Dwyer's studies (only those marked with an asterix in Table 2)



Type of visual (i.e. experimental groups): S = simple line drawing; D = detailed, shaded drawing; M = model photograph; R = realistic photograph. N = non-illustrated version (i.e. control group).

Principal component analysis: the transformation of posttest scores into factor scores

Dwyer obtains at least five measures in each study, i.e. four sets of results from the identification, drawing, terminology and comprehension tests and an overall score which is calculated by combining the results of the aforementioned tests. Such an analysis is rather laborious to perform and prevents reaching an adequate synthesis of results. As can be seen in Table 4, indeed, there are significant correlations between the four posttests, ranging from r = .73 to r = .92. It can also be seen that the comprehension test is not as closely related to the three other tests, which means that it probes some other type of knowledge.

BEST COPY AVAILABLE

Table 4

Pearson's correlation between the four posttests calculated on 159 groups (all correlations $p < 0.0001$)

	Terminology	Drawing	Comprehension
Identification	0.92	0.89	0.76
Terminology	-	0.82	0.79
Drawing	-	-	0.73

These facts lead us to reject working with the four measures and their sum: it appears more practical and useful to work with the two composite variables identified by the factorial method of principal components. In doing so, we are able to re-classify the 159 groups in relation to two factors and then evaluate the various basic hypotheses with the help of analyses of variance. In other words, two different factors replace the original four measures and their sum.

Table 5 shows the principal components derived and rotated with the Varimax Method, from correlations drawn between the drawing, identification, terminology and comprehension tests which have allowed us to identify the two principal factors behind them.

Table 5

Principal components extracted from correlations between the four tests rotated with the Varimax Method (159 groups)

Test	Factor 1	Factor 2
Identification	0.88	0.43
Drawing	0.88	0.38
Terminology	0.79	0.53
Comprehension	0.43	0.90

The first factor explained 59% of the total variance. The phenomena underlying Factor 1 are related above all to the Identification ($r = 0.88$), Drawing ($r = 0,88$), and Terminology tests ($r = 0.79$). Upon examination of the particularities of these tests (see Discussion), we will call it the **Vocabulary factor**. The second factor presented in Table 5 can be used to explain 35% of the total variance, which makes it somewhat less important in the explanation of the total variance. This factor is principally linked to the comprehension test ($r = .90$) and, to a lesser extent, to the terminology test ($r = .53$). We will call it the text-comprehension factor or, more briefly, the **Comprehension factor**.

Factor 1 and Factor 2 scores represent the starting point of our meta-analytic approach and allow us to verify the various hypotheses raised in Dwyer's studies.

Dwyer's total score is actually a composite variable obtained by using the following formula:

$$\text{Total score} = 1*Y_1+1*Y_2+1*Y_3+1*Y_4$$

where Y1 is the raw score of the identification test;
Y_2 is the raw score of the drawing test;
Y_3 is the raw score of the terminology test;
Y_4 is the raw score of the comprehension test.

Factorial scores are calculated by principal components:

$$\text{Factor 1 score} = 0.57419*Y_1+0.64300*Y_2+0.30990*Y_3-0.72669*Y_4$$
$$\text{Factor 2 score} = -0.31033*Y_1-0.42017*Y_2+0.04157*Y_3+1.41281*Y_4$$

where Y_1 is the standardized score of the identification test;
Y_2 is the standardized score of the drawing test;
Y_3 is the standardized score of the terminology test;
Y_4 is the standardized score of the comprehension test.

One of the advantages of the factor scores, as compared to Dwyer's global scores, is that they allow the results of all four tests to be expressed in the same unit of measure. As previously noted, Dwyer adds drawing and comprehension scores, which can reach maximums of 18, to identification and terminology scores, for which maximum scores of 20 can be attained. However, the main advantage of factor scores stems from the distinction that they allow to be made between the contribution of Factor 1, the Vocabulary factor, and Factor 2, the Comprehension factor.

In order to compensate for varying experimental conditions, we re-divided the 159 groups into four different categories:

1. secondary school students - immediate tests;

2. secondary school students - delayed tests;

3. university students - written text;

4. university students - oral text.

In Table 6, the original scores from the four posttests as well as the factor scores are presented separately for each of the four categories.

Table 6

Descriptive statistics of the four posttest scores and the two factor scores presented separately for each category (secondary - immediate test; secondary - delayed test; university - oral text, university - written text)

Measure	Degree	Condition	Mean	SD	N
Drawing	secondary	delayed	8.20	1.87	36
		immediate	9.32	2.44	36
	university	oral	12.10	1.70	47
		written	12.65	2.13	49
Identification	secondary	delayed	8.85	1.48	36
		immediate	9.60	1.73	36
	university	oral	13.19	1.68	47
		written	14.27	1.72	49
Terminology	secondary	delayed	7.86	1.86	36
		immediate	8.12	1.95	36
	university	oral	12.30	1.55	47
		written	14.74	2.07	40
Comprehension	secondary	delayed	7.13	1.22	36
		immediate	11.06	1.85	36
	university	oral	11.07	1.61	47
		written	13.42	1.93	40
Factor 1	secondary	delayed	-0.49	0.62	36
		immediate	-1.09	0.79	36
	university	oral	0.67	0.58	47
		written	0.63	0.70	40
Factor 2	secondary	delayed	-1.20	0.44	36
		immediate	0.56	0.74	36
	university	oral	-0.20	0.65	47
		written	0.82	0.72	40

Generally speaking, secondary school students have results inferior to those of university students, except on the immediate comprehension test in which their results are equal to those of university students in the listening mode. This distinction between secondary school students and university students can be seen in the factor space of Figure 1: secondary school statistics are recorded in the left-hand most portion of Factor 1 and those for university students, to the right.

insert Figure 1

At the secondary level, the delay effect (immediate versus delayed testing) can be detected in the lesser results on three of the tests: comprehension, drawing and identification. The differences can be observed in the factor space for Factor 2 where delayed-test results are recorded in the negative part of the Factor. The narrower margin separating the four delayed-test results explains why these groups' results are almost at the center of the

BEST COPY AVAILABLE

Pictorial factor whereas they are clearly situated in the negative portion of the factor with the immediate-tests' results. Immediate comprehension test results are higher than other immediate test results, and comparable to the university level test results, whereas delayed-comprehension test results compare unfavourably to other test results, which is difficult to understand.

The text presentation mode influences test results at the university level: the written text gives higher test scores for all the tests, and in particular in the comprehension and terminology tests. The presentation-mode effect can also be seen in the factor space of Figure 1: results associated with written text are located in the negative portion of Factor 2 whereas those associated with reading are located in the positive portion of the same factor.

Non-illustrated versus illustrated text versions

The comparisons between the non-illustrated text version (control group) and the illustrated text versions as a whole (experimental groups) through the analysis of the Vocabulary factor (Factor 1) and the Comprehension factor (Factor 2) indicate the following overall results:

Vocabulary (Factor 1): The difference between the non-illustrated test version, on the one hand, and the illustrated test versions, on the other, reveals a highly significant contrast both at the secondary school level ($F = 20.63$, $df = 1$ and 75, $p = .0001$) where it accounts for 17.3% of the total variation of Factor 1 and at the university level ($F = 21.60$, $df = 1$ and 82, $p = .0001$) where it accounts for 21.7% of the total variation. At both levels, the illustrated test versions produce better overall results than those obtained with the non-illustrated text version (see Tables 8 and 10). Clearly, the presence of illustrations has a positive effect on subjects' learning of vocabulary.

Text comprehension (Factor 2): The difference between the non-illustrated and illustrated test versions is nonsignificant at the secondary school level ($F = 0.06$, $df = 1$ and 75, $p = .8143$) and at the university level ($F = 3.59$, $df = 1$ and 32, $p = .0617$). This is to say that Dwyer's illustrations do not have significant impact on the subjects' comprehension of text.

Comparisons of illustrated text versions (Pictorial realism and Color)

In this section the various illustrated Text versions will be subjected to comparison.

Vocabulary (Factor 1): At the secondary level, the global model is significant ($F = 1.95$, $df = 12$ and 51, $p = .0115$). It explains 37% of the total variation. Of the two variables of principal concern, Pictorial realism and Color, only the former is significant (see Table 7). On the other hand, the Delay variable has a highly significant influence on the Vocabulary factor and accounts for 17% of test score variability: after a one-month delay, test score results decline.

BEST COPY AVAILABLE

Table 7

A three-way analysis of variance (ANOVA) with Factor 1 (i.e. vocabulary) as dependent variable: secondary level subjects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Pictorial realism	3	3.305	1.102	3.23	0.0299
Color	1	0.363	0.363	1.06	0.3072
Delay	1	4.787	4.787	14.03	0.0005
Delay x Pictorial realism	3	0.402	0.134	0.39	0.7588
Delay x Color	1	0.017	0.017	0.05	0.8228
Pictorial realism x Color	3	1.374	0.458	1.34	0.2708
Corrected total	63	27.644			

At the secondary school level, the pattern of improvement of test scores in relation to the degree of realism is somewhat cloudy. Looking at Table 8, it can be seen that the best results are associated with detailed drawings and that the worst results are attributed to the realistic photographs. Between these two extremes, the marginal differences obtained for the different degrees of realism do not allow us to determine a ranking for success.

Table 8

Comparison of illustrated and non-illustrated text versions (LSD, multiple T) with Factor 1 (i.e. vocabulary) as dependent variable: secondary level subjects

T Grouping*		Mean	N	VISUAL
	A	-0.350	16	detailed
B	A	-0.541	16	simple
B	A	-0.857	16	model
B		-0.900	16	realistic
	C	-1.794	8	absent

* Different letters identify significant differences

At the university level, the general model (incorporating the oral and written mode of text presentation) is highly significant ($F = 3.30$, $df = 12$ and 65, $p = .0009$). It accounts for 38% of score variation.

Table 9

Three-way analysis of variance (ANOVA) with Factor 1 (i.e. vocabulary) as dependent variable: university level subjects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Pictorial realism	3	6.143	2.048	9.56	0.0001
Color	1	1.937	1.937	9.04	0.0038
Mode	1	0.095	0.095	0.44	0.5079
Mode x Pictorial realism	3	0.334	0.111	0.52	0.6708
Mode x Color	1	0.041	0.041	0.19	0.6641
Pictorial realism x Color	3	0.157	0.052	0.24	0.8655
Corrected total	77	22.406			

BEST COPY AVAILABLE

Factor 1 scores are, at the university level, subject to a highly significant influence from Pictorial realism. This factor accounts for 27% of the variability. As can be seen in Table 10, the simplified drawings lead more to success. Next are the precise drawings and model photographs, which give statistically comparable results. As for the realistic photographs, they result in the poorest test score results among the illustrated text versions.

Table 10

Comparison of Illustrated and Non-illustrated versions (LSD, multiple T) with Factor 1 (i.e. vocabulary) as dependent variable: university level subjects (different letters identify significant differences)

T Grouping*	Mean	N	VISUAL
A	1.126	20	simple
B	0.753	20	detailed
B	0.723	18	model
C	0.359	20	realistic
D	-0.118	9	absent

* Different letters identify significant differences

Color also has a substantial effect on vocabulary learning, accounting for 9% of the variability of this factor. As you can see in Figure 1, color illustrations produce higher recall test scores (M = 0.930) than black and white illustrations (M = 0.609).

Text Comprehension (Factor 2): At the secondary school level, the general model shows a highly significant effect (F = 9.37, df = 12 and 51, p = .0001). This effect, however, must be attributed to the Delay variable, and not to Pictorial realism or Color variables which concern us most (see Table 11). Delay accounts for 65% of the variability whereas Pictorial realism and Color account for no more than 2%.

Table 11

Three-way analysis of variance (ANOVA) with Factor 2 (i.e. text comprehension) as dependent variable: secondary level subjects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Pictorial realism	3	0.845	0.282	0.62	0.6053
Color	1	0.006	0.006	0.01	0.9056
Delay	1	48.559	48.559	106.93	0.0001
Delay x Pictorial realism	3	1.152	0.384	0.85	0.4753
Delay x Color	1	0.014	0.014	0.03	0.8635
Pictorial realism x Color	3	0.478	0.159	0.35	0.7888
Corrected total	63	74.213			

At the university level, the general model also exhibits a highly significant influence on text comprehension (F = 5.61, df = 12 and 65, p = .0001) and accounts for 50% of its variability. As with the Secondary School level, however, the effect is to be attributed to the text-presentation mode rather than to the pictorial variables which concern us most (see Table 12). The mode of presentation has a highly significant effect and accounts for 34% of total

score variability. Written text (s = 0.773) is markedly better-understood than oral text (s = -0.249). Pictorial realism and Color account for a mere 4.25% of total score variability.

Table 12

Three-way analysis of variance (ANOVA) with Factor 2 (i.e. text comprehension) as dependent variable: University level subjects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Pictorial realism	3	1.266	0.422	1.00	0.3969
Color	1	1.224	1.224	2.91	0.0929
Mode	1	18.953	18.953	45.06	0.0001
Mode x Pictorial realism	3	2.485	0.828	1.97	0.1273
Mode x Color	1	0.834	0.834	1.98	0.1638
Pictorial realism x Color	3	2.734	0.911	2.17	0.1004
Corrected total	77	55.662			

Discussion

As has been shown, it is possible to re-analyze Dwyer's results by means of principal components analysis, allowing us to reduce his posttest results to two factors accounting, respectively, for 59% (Factor 1) and 35% (Factor 2) of the total variance shared. The first factor is most closely associated with the identification, drawing and terminology tests whereas the second factor is most closely associated with the comprehension test and, to a lesser degree, with the terminology test. The crucial question is then knowing how these two factors should be interpreted. In other words, the moment has come for us to justify our practice concerning paraphrasing Factor 1 as Vocabulary learning factor and Factor 2 as Text-comprehension factor.

A qualitative-hermeneutic analysis of the characteristics shared (or not shared) by the four posttests is the key to this problem. This means that, with regards to Factor 1, it is a matter of discovering the characteristic that the identification, drawing and terminology tests all have in common and which is absent (or almost absent) from the comprehension test. As for Factor 2, it is a matter of identifying a characteristic inherent in the comprehension test - and to a lesser degree in the terminology test - which is absent (or almost absent) from the drawing and identification tests. Table 13 provides a partial answer to this question. For a randomly selected number of terms used in the four posttests, it indicates the number of times each term is used in the three multiple-choice posttests as a target or as a distractor item (see also Appendices C - E). The average number of words per question is also indicated as a potentially useful indicator: upper-level questions (i.e. text comprehension) should be longer than lower-level questions (i.e. vocabulary).

Table 13

Occurrences of terms(randomly selected) used as target items or distractor items in the four posttests

		MULTIPLE CHOICE		
TERM	DRAWING (target)	IDENTIFICATION (target and distractors)	TERMINOLOGY (target and distractors)	COMPREHENSION (target and distractors)
apex	yes	2	2	-
endocardium	yes	7	6	-
epicardium	yes	4	4	-
left ventricle	yes	3	3	-
myocardium	yes	7	6	-
pericardium	no	7	3	-
right ventricle	no	4	2	-
septum	yes	4	7	-
WORDS* PER QUESTION	1.8	1.7	1.7	3.2

* graphic sequence delimited by blanks (e.g., "left ventricle" = 2 words).

As indicated by the number of occurrences, a given term is proposed as a possible answer in each of the two multiple-choice tests, the identification and the terminology tests - but not in Dwyer's comprehension test. They draw

BEST COPY AVAILABLE

essentially upon the same factual knowledge. Even if Table 13 does not present an exhaustive list, one can see the conceptual proximity of the identification test and the terminology test: response terms used in one of them are also used in the other, and approximately the same number of times. The fact that none of these response terms are present in the comprehension test as well as the average number of words per response-item both indicate well the existing dichotomy we tentatively termed *Vocabulary* and *Text Comprehension*. The average number of words in the comprehension test, which is twice as many as the average number in the two other multiple-choice tests, would even be more if there were not some vocabulary-like questions among the twenty questions (as an example see response item 1 in Appendix E).

With regards to the drawing test (see Appendix B) we believe that, conceptually speaking, it is a vocabulary test too: its target vocabulary is almost identical to that used in the identification and terminology tests.

Both correlational series in Table 5 reflect this observed general dichotomy: Factor 1 is strongly correlated to Dwyer's identification, drawing and terminology tests and Factor 2, to his comprehension test.

The interpretation of Factor 2 as equivalent to a text-comprehension factor seems to us quite straightforward. We think, in fact, that Factor 2 can be explained by the varying degree of the subjects' use of upper-level verbal processes implied by the four posttests. The identification, terminology and drawing tests, all of which are weakly correlated to Factor 2, emphasize lower-level processes. Dwyer's so-called comprehension test is the only one strongly correlated with Factor 2 ($r = .90$) and for this reason its naming seems to us particularly appropriate: its the only test which implies a deeper semantic processing of larger textual units. Indeed, the comprehension test elicits information in such a way so as that the mere recall of terms and / or their spatial relationship to other terms, while remaining a pre-requisite, will not suffice. Comprehension questions about the *function* of the different parts of the heart imply a more profound understanding of the text than the mere spatial identification of some terms (drawing, identification) or their verbal paraphrase (terminology). In the drawing test, which is least correlated with Factor 2 ($r = .38$), the instruction to draw a picture and to spatially locate a list of terms does not allow for the verification of the subjects' text comprehension or their deeper semantic processing of the terms listed. This is also true for the identification test ($r = .43$): an instruction such as "Arrow number ___ points to the ___" only establishes a meta-linguistic link between the verbal and pictorial parts of each multiple-choice item and does not contain semantically elaborated cues related to the experimental text. The terminology test is a little bit more strongly correlated with Factor 2 ($r = .53$). To give the correct answer to multiple-choice questions such as "_?_ is (are) the thick walled chamber's of the heart", one needs to be able to understand and to paraphrase larger units of text.

With regards to Factor 1, which we have tentatively called the Vocabulary factor, this identification is probably too general and has to be restricted to a certain type of vocabulary -- that is **concrete, pictorially representable vocabulary**. Possibly Factor 1 should even be understood as an even more specific concept, i.e. as a factor which reflects concrete, spatially-representable and **actual pictorially test-represented**

BEST COPY AVAILABLE

vocabulary. Indeed, Factor 1 appears closely related to the fact that three tests out of four combine verbal and pictorial elements in some way or another. Only Dwyer's comprehension test is conceptualized using exclusively verbal elements and without making reference to pictorially-represented information. The drawing test requires the students themselves to produce a drawing of the heart as a preliminary step. They must then spatially locate the targeted vocabulary on the sheet of paper. The pictorial representation of the vocabulary is not externally imposed by the drawing test. The drawing of the heart and the spatial identification of the targeted vocabulary can both take advantage of the pictorial component of the identification test, given their conceptual proximity as shown in Table 13. The same explanation is valuable in the case of the terminology test. Despite its apparent lack of any pictorial elements, the terminology test is closely related to the identification test. As a consequence, the pictorial component of the identification test facilitates the subjects' answers both to the terminology and the drawing test which, for this reason, must be considered as being illustrated tests as well (in the sense of externally imposed pictures). The canonical order of Dwyer's four tests, all presented in the same test booklet, is the following: the drawing, identification, terminology and comprehension tests. With all four tests being part of the same booklet, regressions seem possible: the subjects' respect of the suggested chronological order was not really experimentally controlled. So, given the conceptual proximity of the three tests, we suppose that the presence of a picture in one of them, i.e. the identification test (in which its presence is in fact indispensable), helps subjects to answer also the two other tests. This explanation also fits well with the fact that the correlation coefficient between the terminology test and Factor 1 is somewhat weaker (r=.79) than in the case of the other two tests (both r=.88): the paraphrasing task of the terminology test seems less pictorially-dependent, comparatively speaking.

If our interpretation of Factor 1 as *pictorially-dependent* vocabulary test is correct, the empirical and theoretical relevance of results related to this factor unfortunately is weakened because of the systematic experimental bias introduced by the presence of pictures *in the test itself*. The comparison between the illustrated and nonillustrated text versions being made by means of *illustrated* posttests instead of nonillustrated ones, the subjects having seen previously the illustrated text version are iniquitously privileged. The reason for this is that, cognitively speaking, within-modal comparisons result in better success rates than between-modal comparisons. In numerous experiments, between-modal comparisons (e.g. word - picture) and within-modal comparisons (e.g. word - word, picture - picture) are shown to produce significantly different results, the latter being considered in most cases as the cognitively less demanding task (see reviews of Snodgrass, 1980; Clark & Paivio, 1987; Roediger & Weldon, 1987; Glaser, 1992). According to the "encoding specificity principle" (Tulving & Thomson, 1973), recall of pictorial items is favored by the presentation of pictorial test-items and recall of verbal items, by the presentation of verbal test-items. So, in Dwyer's studies the superiority of the illustrated-text subjects' vocabulary learning scores could simply reflect a text-test interactional effect, the word-picture-comparison task being a more demanding task than a picture-picture-comparison task. Because of the presence of a picture in the test itself, the illustrated-text subjects could sometimes use picture -picture comparisons whereas the non-illustrated-text subjects were obliged to use only text-picture comparisons.

BEST COPY AVAILABLE

To conclude this point: In order to demonstrate unequivocally the superiority of illustrated text versions over the non-illustrated text version, you would have to use non-illustrated tests, not illustrated ones. The information tested should not be picture-dependent either. Only Dwyer's Comprehension test satisfies this condition. As previously shown, when Factor 2, which is highly related to Dwyer's comprehension test, was used as the dependent measure, all pictorial main effects and their interactions become nonsignificant.

Conclusions

With this interpretation of both factors in mind, let us summarize the major conclusions of our meta-analysis.

Conclusions with respect to vocabulary learning (Factor 1)

Conclusion 1a: Adding illustrations to the text has a positive significant impact.

Conclusion 2a: Realistic pictures accompanying the text are significantly less effective than abstract ones.

Conclusion 3a: The use of color in pictures has a significantly beneficial effect with university students but no effect with secondary school students.

Conclusion 4a: The degree of pictorial realism (i.e. simple line drawing, detailed drawing, photograph of a model, realistic photograph) does not interact significantly with the presence or absence of color.

Conclusion 5a: The degree of realism does not interact significantly with the presentation mode of the text (oral, written).

Conclusion 6a: The presence or absence of color does not interact significantly with the presentation mode of text.

Conclusion 7a: The degree of realism does not interact significantly with the test delay (immediate test, delayed test).

Conclusion 8a: The presence or absence of color does not interact significantly with the test delay.

Conclusions with respect to text comprehension (Factor 2)

Conclusion 1b: Adding illustrations to the text has no significant effect.

Conclusion 2b: Realistic and abstract pictures accompanying the text do not give significantly different results.

Conclusion 3b: With secondary school and university students, color has no

significant impact on the presentation mode of the text.

Conclusion 4b: The degree of realism does not interact significantly with test delay (immediate test, delayed test).

Conclusion 5b: The degree of realism does not interact with the presentation mode of text (oral, written).

Conclusion 6b: The presence or absence of color does not interact with the presentation mode of text.

Conclusion 7b: The degree of realism does not interact with test delay (immediate test, delayed test).

Conclusion 8b: The presence or absence of color does not interact with test delay.

Concerning main effects (conclusions 1 - 3), there is a profound dichotomy between Factor 1 and Factor 2 results: the absence and presence of pictures (Conclusion 1a), the degree of pictorial realism (Conclusion 2a) and the absence and presence of color (Conclusion 3a) are all significant variables with respect to the Vocabulary factor, but nonsignificant with respect to the text-comprehension factor (Conclusions 1b, 2b and 3b). Keep in mind, however, that our interpretation of the Vocabulary factor as possibly *pictured*-vocabulary factor limits the relevance and generalizability of the (a)-conclusions.

According to conclusion (2a), Pictorial realism is a significant variable (even if not all of the six pairwise comparisons are significant, which is the case especially in the experiments done with secondary school level students; see Tables 7 and 8). In general, conclusion (2a) seems to corroborate Dwyer's conclusion that the less-realistic visuals are more effective than the realistic ones (e.g., Dwyer, 1972b). This superiority which seems to bear witness to the difficulty subjects experience dealing with an overabundance of pictorial information and as such lends credence to the saying "sometimes more is less". This interpretation may be tentative, however, it could also simply reflect another text-test interactional effect: the visual used in Dwyer's identification test - a detailed, shaded drawing - is identical to the visuals presented in two of the eight experimental text versions (see Table 2, versions 3 and 4) whereas the visuals used in the other illustrated text versions differ, to varying degrees, from this test visual. Methodologically, this experimental bias favors certain groups and puts others at a disadvantage. The test results - particularly those of the university students - could reflect this bias. While they are the least effective, the realistic photographs are also those that differ the most from the visuals used in the identification test (which, as we have argued before, directly influences the other two tests with regards to Factor 1).

Similar caution must be shown with respect to color (conclusion 3a): since the identification test is illustrated in black and white, one could think that the experimental groups having seen the colored text versions may be at a disadvantage as compared to the groups with black and white versions, even if

Lamberski's (1980) and Lamberski's and Dwyer's (1983) experimental data do not support this hypothesis. Both authors explored the relationship between the absence and presence of color during experimental treatment and the test (2 x 2), with nonsignificant interactional effects between both variables. At first glimpse, the lack of interaction seems to be a solid basis for excluding the much-feared experimental bias, yet the experimental material used by Lamberski and Dwyer differs dramatically from the earlier studies: half of the test items have been modified or replaced.

These studies as well as de Melo's (1980) study (a doctoral thesis supervised by Dwyer) clearly show that Dwyer was aware of the possibility of experimentally uncontrolled text-test interactions with respect to the variables mentioned in our conclusions (1a) - (3a). However with regards to the degree of pictorial realism, Dwyer and his collaborators did not perform any analogous experiment varying the abstractional level of the pictures in the text and the identification test.

To summarize, we think that the relevance of the significant main effects stated in conclusions (1a) and (2a) are substantially weakened because of an experimental bias favoring some experimental groups compared to others. Conclusion (3a) could also be affected by this bias. For this reason, Factor 2 gets even more important for Dwyer's central hypotheses concerning pictorial realism. Unfortunately, all Factor-2 results reveal only nonsignificant differences.

Conclusions (4) - (8) all concern experimentally controlled interactional effects. We will treat the results for both dependent measures together as all of them contribute only nonsignificant interactional effects.

Concerning the conclusions (4a) and (4b), the absence of a significant interaction between the degree of realism and color seems all the more revealing in that color was not employed in a similar fashion in all types of visuals. According to the descriptions made by Dwyer (1967c) of the experimental text versions, color plays a *representational* function only in the photograph versions and in the detailed-drawings versions (see Table 1, versions 4 - 9). In the simple-line-drawings versions (versions 2 and 3), however, the choice of color is said to be arbitrary, adding nothing to the pictorial realism: "The line drawings (...) were blue and pink (blue lines on a pink field)", (p. 18). Theoretically, such asymmetrical design should promote the appearance of a significant interaction, which was not the case.

According to conclusions (5) - (8), Pictorial realism and Color do not interact either with the text presentation mode (oral, written) or with the delay type between treatment and test (immediate tests, delayed tests). It would be worthwhile examining these results in the context of studies having explored the possible interaction between the picture variable (absence, presence) and the color variable (color, b & w). Intuitively, we expect the interaction of Delay or Presentation mode with Pictorial realism to be significant only if it is clearly demonstrated that Delay and Presentation mode indeed interact significantly with the more fundamental pictorial variable opposing the absence and the presence of pictures.

The doctoral thesis of Joseph (1978) supervised by Dwyer, as well as articles by Rohwer & Harris (1975), Nugent (1982) and Sewell & Moore (1980) allow for

the exploration of the relationship between the text-Presentation mode (oral, written) and the presence or absence of pictures. With the exception of Sewell & Moore (1980), these studies indicate that the presence of illustration is especially effective with written text.

Duchastel's (1981) study has often been brandished as proof of an existing interactional effect between Delay (immediate test, delayed test) and Illustration (absence, presence) in texts. The study reputedly confirms the functional typology of illustration previously developed by the author (Duchastel, 1978): one of the three major functions of the illustrations would be to support the delayed recall of the text that it accompanies. However, we believe that the beneficial influence of illustration in delayed testing, compared to immediate testing, is imputable to inter-group inequalities in Duchastel's experimental design. Contrary to most studies, Duchastel assigns independent groups to immediate and delayed recall; a within-subjects treatment of this variable would avoid this type of inequality. Delay x Illustration interactions are indeed generally nonsignificant (Peeck, 1974; Jahoda, Cheyne et al., 1976; Haring & Fry, 1979; Purkel & Bornstein, 1980; Gilmartin, 1982; Hannafin, 1983; Perrett, 1985; Tajika, Taniguchi, Yamamoto & Mayer, 1988; Smith & Smith, 1991), with two exceptions (Bernard, Petersen & Ally, 1981), but both are incompatible with Duchastel's conclusion.

To summarize, our meta-analysis of Dwyer's experiments shows significant main effects only for the Vocabulary factor, a factor we suppose to reflect in some part an experimental biais. As a consequence, the empirical and theoretical relevance of conclusion (2a) - the main goal of Dwyer's work - must be questioned. Factor 2 as dependent measure is therefore crucial to Dwyer's central hypotheses concerning pictorial realism (including color). Factor 2 is supposed to reflect the subjects' comprehension of text. Unfortunately, all text-comprehension conclusions reveal only nonsignificant differences. For this reason, we think that Dwyer's experiments fail to prove what they purport to prove: our meta-analysis does not confirm or support Dwyer's central hypotheses.

References

AMSDEN, R. H. (1960). Children's preferences in picture story book variables. *Journal of Educational Research*, 53 (8), 309-312.

BERNARD, R. M., PETERSON, C. H. & ALLY, M. (1981). Can images provide contextual support for prose? *Educational Communication and Technology Journal*, 29 (2), 101-108.

BLOOMER, R. H. (1960). Children's preferences and responses as related to styles and themes of illustration. *Elementary School Journal*, 60, 334-340.

BOHLE, R. H. & GARCIA, M. R. (1987). Reader response to color halftones and spot color in newspaper design. *Journalism Quarterly*, 64, 731-739.

BOOTH, G. D. & MILLER, H. R. (1974). Effectiveness of monochrome and color presentations in facilitating affective learning. *AV Communication Review*, 22, 409-422.

BURDICK, J. G. (1959). *A study of cross-section drawings used as technical illustrations in High School science textbooks*. Doctoral Dissertation, Syracuse University.

CHUTE, A. G. (1980). Effect of color and monochrome versions of a film on incidental and task-relevant learning. *Educational Communication & Technology*, 28 (1), 10-18.

CLARK, J. M. & PAIVIO, A. (1987). A dual coding perspective on encoding processes. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes*, pp. 5-33. New York: Springer-Verlag.

DUCHASTEL, P. (1978). Illustrating instructional texts. *Educational Technology*, 18 (11), 36-39.

DUCHASTEL, P. C. (1981). Illustrations in text: a retentional role. *Programmed Learning and Educational Technology*, 18 (1), 11-15.

DE MELO, H. T. (1980). *Visual self-paced instruction and visual testing in biological science at the secondary level*. Doctoral Thesis, Pennsylvania State University.

DENIS, M. & DE POUQUEVILLE, P. (1976). Le réalisme de la figuration dans la mémoire d'actions concrètes. *Bulletin de Psychologie*, 30, 543-550.

DWYER, F. M. (1967a). Adapting visual illustrations for effective learning. *Harvard Educational Review*, 37, 250-263.

DWYER, F. M. (1967b). The relative effectiveness of varied visual illustrations in complementing programmed instruction. *Journal of Experimental Education*, 36 (2), 34-42.

DWYER, F. M. (1967c). *A study of the relative effectiveness of varied visual illustrations*. U.S. Department of Health, Education, and Welfare, (Report

no. BR-6-8840), ERIC Document ED 020658.

DWYER, F. M. (1968a). The effect of visual stimuli on varied learning objectives. *Perceptual and Motor Skills*, 27, 1067-1070.

DWYER, F. M. (1968b). Effect of varying amount of realistic detail in visual illustrations designed to complement programmed instruction. *Perceptual and Motor Skills*, 27, 351-354.

DWYER, F. M. (1968c). The effectiveness of visual illustrations used to complement programed instruction. *Journal of Psychology*, 70, 157-162.

DWYER, F. M. (1968d). An experiment in visual learning at the eleventh-grade level. *Journal of Experimental Education*, 37 (2), 1-6.

DWYER, F. M. (1968e). When visuals are not the message. *Educational Broadcasting Review*, 2, 38-43.

DWYER, F. M. (1969a). An experiment in visual communication. *Journal of Research in Science Teaching*, 6, 185-195.

DWYER, F. M. (1969b). Student perception of the instructional value of visual illustration. *Medical and Biological Illustration*, 19, 42-45.

DWYER, F. M. (1970a). The effect of image size on visual learning. *The Journal of Experimental Education*, 39 (1), 36-41.

DWYER, F. M. (1970b). Exploratory studies in the effectiveness of visual illustrations. *AV Communication Review*, 18, 235-249.

DWYER, F. M. (1970c). Effect of questions on visual learning. *Perception and Motor Skills*, 30, 51-54.

DWYER, F. M. (1971a). Assessing student's perceptions of the instructional value of visual illustrations used to complement programmed instruction. *Programmed Learning and Educational Technology*, 8 (2), 73-80.

DWYER, F. M. (1971b). Color as an instructional variable. *AV Communication Review*, 19 (4), 399-416.

DWYER, F. M. (1971c). Effect of questions on visualized instruction. *Journal of Psychology*, 78, 181-183.

DWYER, F. M. (1971d). An experimental evaluation of the instructional effectiveness of black-and-white and colored illustrations. *Didakta Medica*, 3 (4), 96-101.

DWYER, F. M. (1971e). Questions as advanced organizers in visualized instruction. *Journal of Psychology*, 78, 261-264.

DWYER, F. M. (1971f). Student perceptions of the instructional effectiveness of black & white and colored illustrations. *Journal of Experimental*

Education, 40 (1), 28-34.

DWYER, F. M. (1972a). The effect of overt responses in improving visually programed science instruction. *Journal of Research in Science Teaching*, 9, 47-55.

DWYER, F. M. (1972b). *A guide for improving visualized instruction*. State College, Pennsylvania: Learning Services.

DWYER, F. M. (1975). On visualized instruction effect of students' entering behavior. *Journal of Experimental Education*, 43 (3), 78-83.

DWYER, F. M. (1976). The effect of IQ level on the instructional effectiveness of black-and-white and color illustrations. *AV Communication Review*, 24 (1), 49-62.

DWYER, F. M. (1978). *Strategies for improving visual learning*. State College, Pennsylvania: Learning Services.

DWYER, F. M. & LAMBERSKI, R. J. (1982). A review of the research on the effects of the use of color in the teaching-learning process. *International Journal of Instructional Media*, 10 (4), 303-328.

FLEMING, M. L. & SHEIKHIAN, M. (1972). Influence of pictorial attributes on recognition memory. *AV Communication Review*, 20 (4), 423-441.

FRANZWA, D. (1973). Influence of meaningfulness, picture detail, and presentation mode on visual retention. *AV Communication Review*, 21 (2), 209-223.

GATTEGNO, C. (1966). *La lecture en couleurs - guide du maître*. Neuchâtel: Delachaux & Niestlé.

GIASSON-LACHANCE, J. & LEDUC, A. (1981). L'utilisation de la couleur dans la mise en relief de graphèmes chez des enfants de classes de maternelle. *Rééducation Orthophonique*, 79, 429-439.

GILMARTIN, P. P. (1982). The instructional efficacy of maps in geographic text. *Journal of Geography*, 81, 145-150.

GLASER, M. O. & GLASER, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, 8 (6), 875-894.

GLASER, W. R. (1992). Picture naming. *Cognition*, 42, 61-105.

GOLDSMITH, E. (1984). *Research into illustration: An approach and a review*. Cambridge University Press.

HANNAFIN, M. J. (1983). The effects of instructional stimulus loading on the recall of abstract and concrete prose. *Educational Communication and Technology Journal*, 31 (2), 103-109.

BEST COPY AVAILABLE

HARING, M. J. & FRY, M. A. (1979). Effect of pictures on children's comprehension of written text. *Educational Communication and Technology Journal*, 27 (3), 185-190.

HAYES, D. A. & HENK, W. A. (1986). Understanding and remembering complex prose augmented by analogic and pictorial illustration. *Journal of Reading Behavior*, 28 (1), 63-78.

HINDS, L. (1966). *An evaluation of words in color or morphologico- algebraic approach to teaching reading to functionally illiterate adults.* Case Western Reserve University.

HOUGHTON, H. A. & WILLOWS, D. M. (1987). *The psychology of illustration. Volume 2: Instructional issues.* New York: Springer.

HURT, J. A. (1987). Assessing functional effectiveness of pictorial representations used in text. *Educational Communication and Technology Journal*, 35 (2), 85-94.

JAHODA, G., CHEYNE, W. M., DEREGOWSKI, J. B., SINHA, D. & COLLINGBOURNE, R. (1976). Utilization of pictorial information in classroom learning: A cross cultural study. *AV Communication Review*, 24 (3), 295-315.

JAGODZINSKA, M. (1976). The role of illustrations in verbal learning. *Polish Psychological Bulletin*, 7 (2), 95-104.

JOSEPH, J. H. (1978). *The instructional effectiveness of integrating abstract and realistic visualization.* Doctoral thesis, Pennsylvania State University.

KATZMAN, N. & NYENHUIS, J. (1972). Color vs black-and-white effects on learning, opinion, and attention. *AV Communication Review*, 20 (1), 16-28.

KOENKE, K. & OTTO, W. (1969). Contribution of pictures to children's comprehension of the main idea in reading. *Psychology in the Schools*, 6, 298-302.

KOSKI, G. R. (1975). *A study to examine the effects on cognitive learning and attitude of two extreme illustration production costs in a slide-tape program.* Doctoral thesis, Michigan State University.

LAMBERSKI, R. J. (1980). *The effect of a verbal and visual color code on self-paced instruction and testing for retention on different tasks.* Doctoral thesis, Pennsylvania State University.

LAMBERSKI, R. J. & DWYER, F. M. (1983). The instructional effect of coding (color and black and white) on information acquisition and retrieval. *Educational Communication and Technology Journal*, 31 (1), 9-21.

LAMBERSKI, R. J. & ROBERTS, D. M. (1979). *Efficiency of students' achievement using black, white and color coded learning and test materials.* ERIC Document ED 172 800.

LEVIE, W. H. & LENTZ, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology*, 30 (4), 195-233.

MCALISTER, B. K. (1991). *Effects of analogical vs. schematic illustrations on initial learning and retention of technical material*. Eric Document ED 341 838.

NUGENT, G. C. (1982). Pictures, audio, and print: Symbolic representation and effect on learning. *Educational Communication and Technology Journal*, 30 (3), 163-174.

PARKHURST, P. E. (1974). *Assessing the effectiveness of self-paced visualized instruction; a multifactor analysis on five different educational tasks*. Doctoral Thesis, Pennsylvania State University.

PARKHURST, P. E. & DWYER, F. M. (1983). An experimental assessment of students' IQ level and their ability to benefit from visualized instruction. *Journal of Instructional Psychology*, 10, 9-20.

PEECK, J. (1974). Retention of pictorial and verbal content of a text with illustrations. *Journal of Educational Psychology*, 66 (6), 880-888.

PERRETT, R. D. J. (1985) *The effects of text illustrations on the memorability of text for students differing in reading ability*. Doctoral Thesis, Indiana University.

PURKEL, W. & BORNSTEIN, M. H. (1980). Pictures and imagery both enhance children's short-term and long-term recall. *Developmental Psychology*, 16 (2), 153-154.

RAMSEY, I. L. (1982). Effect of art style on children's picture preferences. *Journal of Educational Research*, 75 (4), 237-240.

READENCE, J. E. & MOORE, D. W. (1981). A meta-analytic review of the effect of adjunct pictures on reading comprehension. *Psychology in the Schools*, 18, 218-224.

REID, D. J., BRIGGS, N. & BEVERIDGE, M. (1983). The effect of picture upon the readability of a school science topic. *British Journal of educational Psychology*, 53, 327-335.

ROEDIGER, H. L. & WELDON, M. S. (1987). Reversing the picture superiority effect. In M. A. McDaniel & M. Pressley (eds.), *Imagery and related mnemonic processes*, p. 151-176. New York: Springer.

ROHWER Jr., W. D. & HARRIS, W. J. (1975). Media effects on prose learning in two populations of children. *Journal of Educational Psychology*, 67 (5), 651-657.

SAMUELS, S. J., BIESBROCK, E. & TERRY, P. R. (1974). The effect of pictures on children's attitudes toward presented stories. *Journal of Educational Research*, 67 (6), 243-246.

SEWELL Jr., E. H. & MOORE, R. L. (1980). Cartoon embellishments in informative presentations. *Educational Communication and Technology Journal*, 28 (1), 39-46.

SLOAN, M. A. (1971). *Picture preferences of elementary school children and teachers*. Doctoral Thesis, State University of New York at Buffalo.

SMITH, M. A. & SMITH, P. L. (1991). *Effects on concretely versus abstractly illustrated instruction on learning abstract concepts*. ERIC Document ED 335 014.

SNODGRASS, J. G. (1980). Towards a model for picture-word processing. In P. A. Kolers, M. E. Wrolstad & H. Bouma (Eds.), *Processing of visible language* (Vol. 2, pp. 565-584). New York: Plenum Pres.

STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 404-426.

TABER, J. I. & GLASER, R. (1962). An exploratory evaluation of a discrimination transfer learning program using literal prompts. *The Journal of Educational Research*, 55 (9), 508-512.

TAJIKA, H., TANIGUCHI, A., YAMAMOTO, K. & MAYER, R. E. (1988). Effects of pictorial advance organizers on passage retention. *Contemporary Educational Psychology*, 13, 133-139.

THOMAS, J. L. (1978). The influence of pictorial illustrations with written text and previous achievement on the reading comprehension of fourth grade science students. *Journal of Research in Science Teaching*, 15 (5), 401-405.

Tulving, E. & Thomson, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373.

WADDILL, P. J., MCDANIEL, M. A. & EINSTEIN, G. O. (1988). Illustrations as adjuncts to prose: A text-appropriate processing approach. *Journal of Educational Psychology*, 80 (4), 457-464.

WHEELBARGER, J. J. (1970). *An investigation of the role of pictorial complexity in visual perception*. Doctoral Thesis, University of Viginia, Eric Document ED 044 038.

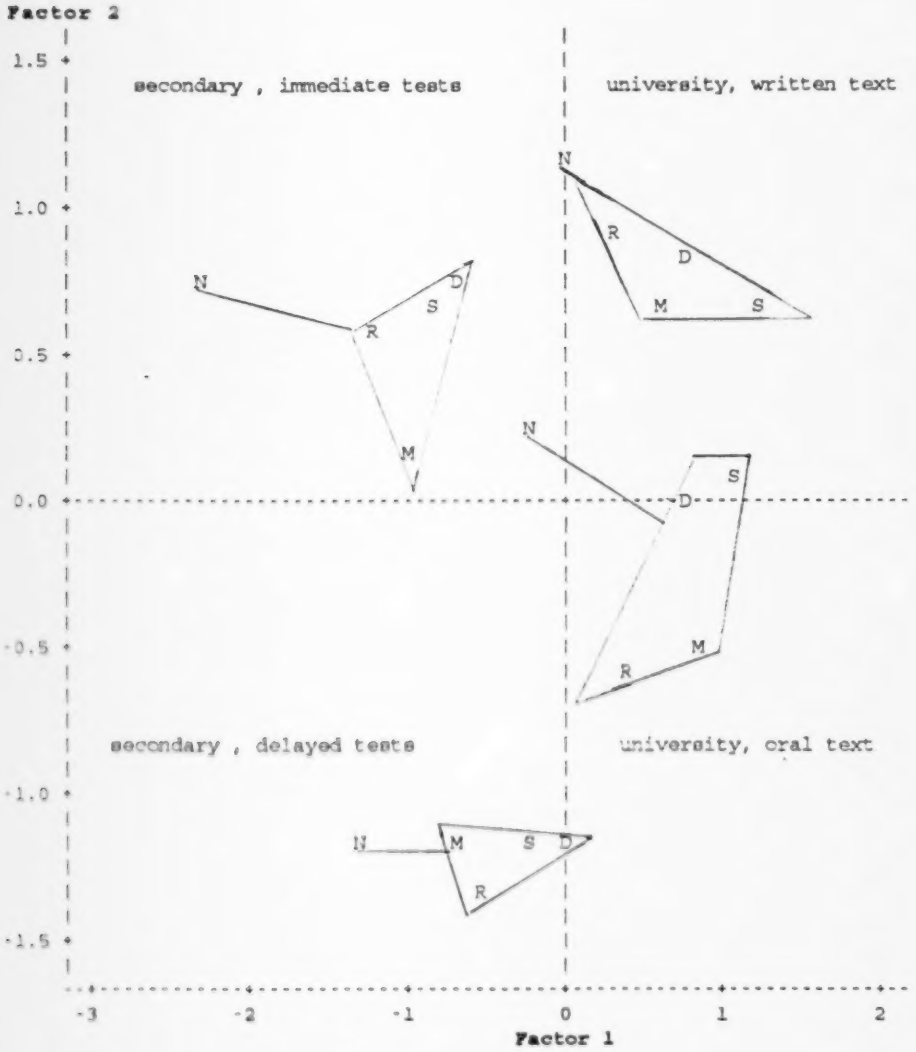
WILLOWS, D. M. & HOUGHTON, H. A. (1987). *The psychology of illustration. Volume 1: Basic research*. New York: Springer.

NOTE 1

REINWEIN, J. (to be published in 1997). *Au carrefour de l'illustration - Vol. I: L'illustration et le texte. Bibliographie analytique des recherches expérimentales.*
(Approximately 650 pages. Volume I contains 327 picto-verbal experiments on sentence and text level described according to the most important parameters in picto-verbal research.)

Figure 1

Factorial classification of Pictorial Realism
(S = simple line drawing; D = detailed, shaded drawing; M = model photograph;
R = realistic photograph; N = nonillustrated)



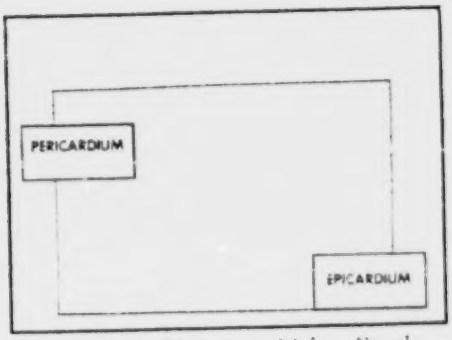
APPENDIX B

Drawing test

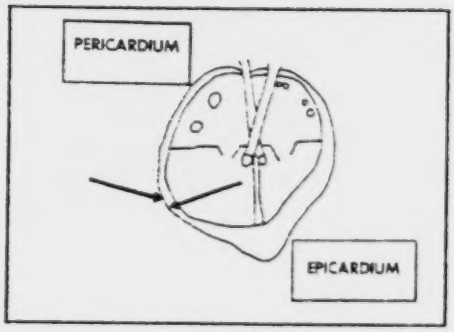
Draw a picture of a heart and place the number of the identified parts where they would be located on the heart.

1. superior vena cava
2. aorta
3. tricuspid valve
4. pulmonary vein
5. septum
6. epicardium
7. aortic valve
8. pulmonary valve
9. inferior vena cava
10. pulmonary artery
11. myocardium
12. endocardium
13. mitral valve
14. right auricle
15. right ventricle
16. left auricle
17. left ventricle
18. apex

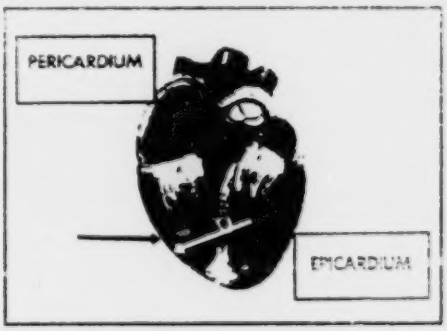
APPENDIX A
Sample of visuals



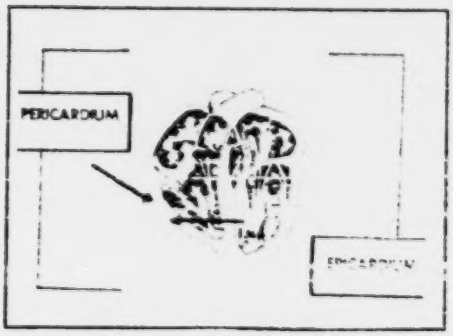
Group 1: Oral Instruction Without Visuals



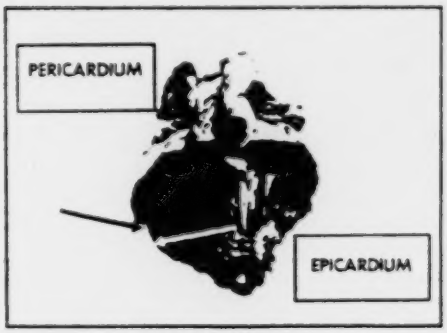
Group 2: Simple Line Presentation (b-w)



Group 6: Heart Model Presentation (b-w)



Group 4: Detailed, Shaded Drawing Presentation (b-w)



Group 8: Realistic Photographic Presentation (b-w)

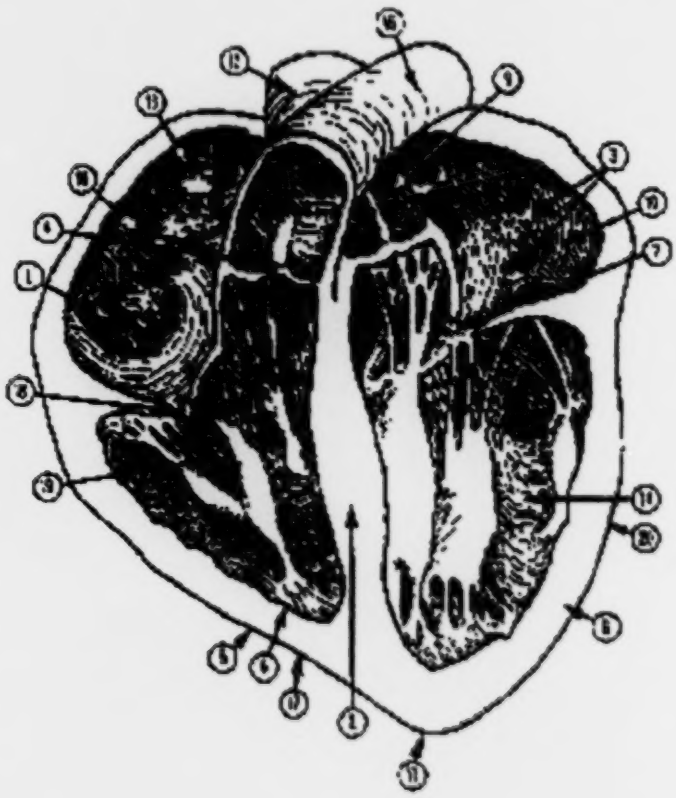
BEST COPY AVAILABLE

APPENDIX C

A multiple-choice item of the Identification test (example)

Select the answer you feel best identifies the part of the heart indicated by the numbered arrows and mark the corresponding letter on the provided answer sheet.

2. Arow number two (2) points to the
   a. pericardium
   b. endocardium
   c. septum
   d. myocardium
   e. pulmonary artery



BEST COPY AVAILABLE

APPENDIX D

A multiple-choice item of the Terminology test (example)

This section requires that you select the answer that best completes the sentence. Mark the correct letter on the provided answer sheet.

10. The _____ is another name for the part of the heart called the heart muscle.

a. apex
b. epicardium
c. endocardium
d. myocardium
e. septum

APPENDIX E

A multiple-choice item of the Comprehension test (2 examples)

In the following multiple-choice questions, select the answer which you feel best answers the question, and place the corresponding letter on you answer sheet.

1. Which valve is most like the tricuspid in function?
   a. pulmonary
   b. aortic
   c. mitral
   d. superior vena cava

19. If the aortic valve is completely closed which of the following statements is correct. The
   a. systolic phase of the heart is occurring
   b. diastolic phase of the heart is occurring
   c. mitral and tricuspid valves are completely shut
   d. blood is rushing into the right and left auricles